VIGILANTE: ULTRAFAST SMART SENSOR FOR TARGET RECOGNITION AND PRECISION TRACKING IN A SIMULATED CMD SCENARIO

Suraphol Udomkesmalee
Jet Propulsion Laboratory
Pasadena, California

Steven C. Suddarth
Ballistic Missile Defense Organization
Washington, DC

ABSTRACT

VIGILANTE is an ultrafast smart sensor testbed for generic Automatic Target Recognition (ATR) applications with a series of capability demonstration focussed on cruise missile defense (CMD). VIGILANTE's sensor/processor architecture is based on next-generation UV/visible/IR sensors and a tera-operations per second sugar-cube processor, as well as supporting airborne vehicle. Excellent results of efficient ATR methodologies that use an eigenvectors/neural network combination and feature-based precision tracking have been demonstrated in the laboratory environment.

INTRODUCTION

The past 30 years have witnessed tremendous improvements in sensors and processors for solving the Automatic Target Recognition (ATR) problem [1]. Recent advances in microprocessor and parallel-processor hardware provide hope for automated scene interpretation that mimics human visual intelligence. Since the first 0.06 MIPS computer-on-a-chip introduced by Intel in 1971, today's microprocessors have grown to contain many millions of transistors and crunch numbers approaching 1000 MIPS [2]. Despite of these advances, we continue to struggle with design of a deployable system for ATR in real-time.

VIGILANTE is a new sensing/processing architecture [11] comprised of the next-generation UV/visible/IR sensors and a high-speed, low-power sugarcube-sized processor and offers hope of achieving a robust, real-time ATR system in a small package. Using the core computing engine developed under the BMDO 3-dimensional artificial neural network (3DANN) program [12]-[13], 64 parallel convolution operations using up to a 64x64 kernel size can be carried out at tera-operations per second (10^{12} OPS). This new processing possibility creates a larger set of feature images from one raw image and fuses this set of data to arrive at the final interpretation of the scene in real-time; this is unthinkable in conventional digital and optical processing medium. Excellent results have already been demonstrated using a hierarchy of eigenvector templates and 3-layer feedforward neural network classifier to fuse these projections of the input scene into 64 eigenvector components [14]. Experiments with a geometrically constrained feature-based algorithm for precision tracking also yield promising results.

The sensors are the Quantum Well Infrared Photodetector (QWIP), the Active Pixel Sensor (APS), and the delta-doped ultraviolet charge-coupled device (UV CCD). These three sensors cover the wavelength ranges 8 to 9, 0.5 to 0.9 and 0.3 to 0.7 μm, respectively. VIGILANTE's sensors can be queued to assist in the ATR functions of detection, classification, and precision tracking. For example, UV wavelengths (0.3 to 0.7 μm) can be used for plume detection, IR (8 to 9 μm) is suitable for cold-body sensing/ classification, and the visible wavelengths (0.5 to 0.9 μm) can provide close-up tracking for aim-point selection in the end-game. Eventually, the VIGILANTE sensors may be used for simultaneous fusion of the data from all wavelengths.

VIGILANTE will provide a complete multisensor and processing system in a small package that is suitable for ATR and pave the way for unique onboard, real-time processing of sensor images for autonomous interceptors and general-surveillance systems. Real-time target recognition will be demonstrated through a series of ground/airborne experiments using real target images.

WHAT'S UNIQUE ABOUT VIGILANTE

VIGILANTE approaches the target recognition problem from a new angle with specialized processing hardware to do the job.

Computer vision today: Classical techniques generally solved this problem by *template matching*[3]. By creating a template (or mask) of an intensity profile that represents the target and scanning it across the image, the region that was most similar to the previously specified template was then declared "winner." The required scanning process for locating the target anywhere in the scene was computationally cumbersome, requiring as many steps as the product of the number of pixels in the image times the number of pixels in the template.

Optoelectronics designers have tried to speed up the process of cross-correlation (template

matching) by using the fact that correlation can be performed in the frequency domain by multiplying the Fourier transforms of the target image and the template image. In inverse Fourier transform is then performed on the output image to locate the match. Optical correlators provide high-speed template matching [4], and continuing advancements in spatial light modulators and binary phase-only filters [5] have made the packaging of optical correlators much more practical for real-time use. However, because the target can be viewed from many angles, ranges, and lighting conditions, generating and using a family of templates to cover all possible distorted views of the target is computationally intractable, even with the optical processing alternative. To avoid the requirement of cycling through millions of templates for each target, composite filter designs have been proposed [6]. This attempt to synthesize the "silver bullet" template faces many performance and system issues, such as signal-to-noise, discrimination ability, programmability and limitations of available spatial light modulators, post-processing of correlation outputs, and packaging. All of these prevent the realization of a flexible, robust optical processor for ATR.

Computer vision researchers have made some progress by systematically breaking the target recognition task into a lengthy processing-step sequence consisting of scene-enhancement preprocessing, target-background segmentation, geometric/texture feature extraction, and knowledge-based recognition [7]-[8]. Repeated iterations reduce raw-pixel data into smaller sets of feature points for interpretation involving both linear and nonlinear filtering operations. Often, large main frame computers and specialized supercomputer systems are needed here. Today's specialized parallel processor systems [9]-[10] developed for real-time image processing are still limited to small-kernel convolutions and can not meet the processing demands of real-time ATR. Defense systems have focused on high-performance infrared, millimeter wave, and other wavelength sensors with the hope of simplifying ATR processing requirements by reducing background clutters (in addition to being able to "see" through weather and darkness). Finally, assuming that simple architectures can get beyond target detection (and perhaps classification), eventual target recognition and aim-point selection still demand the heavy-computation machinery that is not suitable for deployment.

VIGILANTE's contribution: The VIGILANTE processing architecture makes the most of whatever imagery it is presented, whether monochromatic, multispectral, motion or still through the use of an architecture modelled after the human visual system. The VIGILANTE view of image recognition processes is to break them down into four stages: collection of imagery, generation of synthetic imagery, image fusion, and semantic interpretation. Figure 1 shows a simplified model of the eye-brain system.

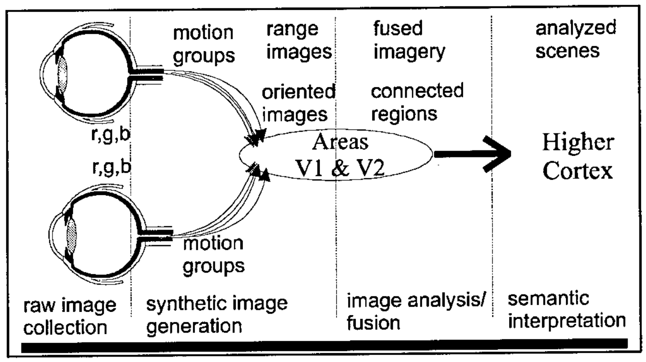


Figure 1: Notional eye-brain process architecture

It could be argued that the brain uses synthetic imagery to analyze scenes by comparing corresponding pixels among the various imagery types. This is essentially a "rich pixel" concept, where the brain becomes a data fusion machine at a pixel level before analyzing the scene in a semantic way. Enriching pixels could be seen as a way of improving evidence used in properly classifying each pixel of the image. Simply put, rich pixel processing consists of the following steps:

1. Generating synthetic images that augment raw images with additional information
2. Fusing all images
3. Interpreting the fused images

VIGILANTE's philosophy, therefore, is to create a machine whose fundamental data type is an entire image plane, rather than more basic data types such as integer or floating point numbers. By breaking down complex image recognition tasks into a series of regular operations, VIGILANTE maps these functions to a relatively small set of special-purpose hardware that can implement a wide variety of algorithms. In VIGILANTE, synthetic image generation tasks (e.g. spatial filtering, motion, and correspondence) use special-purpose hardware, namely the 3DANN-M convolution device. Pixel-level fusion, although less regular than most image

generation tasks can be performed on regular parallel architectures such as SIMD arrays. Semantic analysis seldom presents a significant computational bottleneck compared to the other functions. This task can normally be handled with general-purpose hardware. This mapping concept is illustrated in Figure 2.

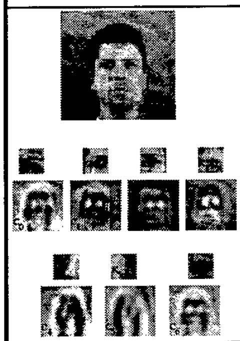
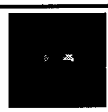
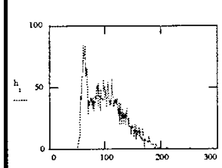
Examples	Type of Process	Optimal Hardware
	N^4 - Convolution - Gray-scale morphology - Rotation/scale invariant patterns	Some special purpose circuits
$P = N(C_1 - C_2)$ M_0  	N-point ops - Summing olding ng	Various massively parallel processors, typically SIMD
 	"Other" - Connected component - Local histograms - Semantic interpretation	Various parallel (MIMD, SIMD) and serial processors based upon algorithm

Figure 2: Classes of processing primitives for VIGILANTE's rich-pixel processing

SYSTEM DESCRIPTION

VIGILANTE consists of the Viewing Imager/Gimballed Instrumentation Laboratory (VIGIL) and Analog Neural Three-dimensional processing Experiment (ANTE). VIGIL is an airborne telescope serving the dual functions of data acquisition for target recognition experiments and testing of novel active and passive focal plane imagers. The telescope will ultimately consist of a self-contained 15-cm Cassegrain unit, a gimbaled mirror, and channels for multiband sensors. A schematic diagram of the VIGILANTE system is shown in Figure 3.

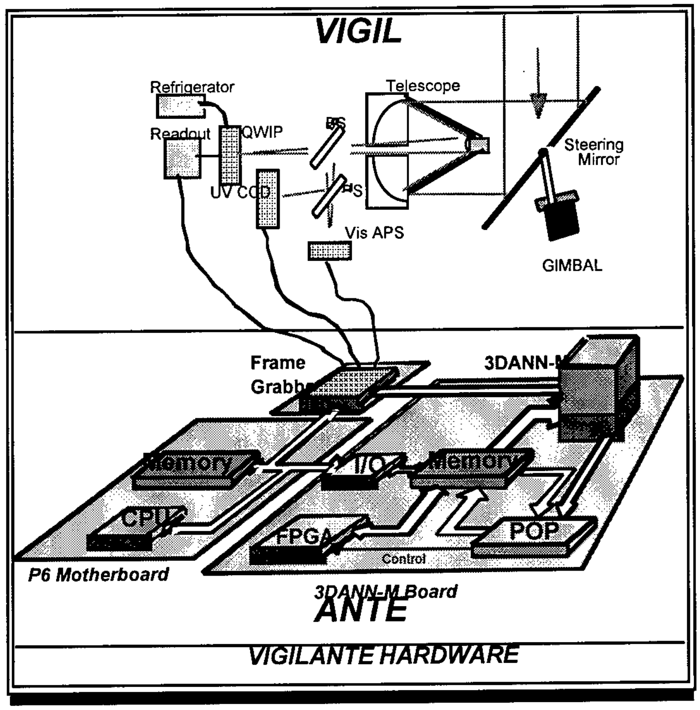


Figure 3: A schematic diagram of the VIGILANTE system. VIGIL is an integrated optical system that splits/transmits the incoming light (steered by a gimbaled mirror) detected by the respective IR/visible/UV sensors. ANTE is the processing system that selects each sensor channel for processing that is done by a commercial frame buffer and host processor and carries out real-time ATR by means of specialized, analog neural networks (3DANN-M) and a point operation processor (POP).

ANTE is a prototype image-processing/target-recognition computer architecture based upon technology developed under the ongoing 3-dimensional artificial neural network (3DANN) program. 3DANN is a sugar-cube-sized, low-power neuroprocessor with its IC stack mated to an IR sensor array. ANTE uses a modified version of the 3DANN referred to as 3DANN-M (or commercially known as TeraCon). The special modifications to 3DANN allow VIGILANTE to accept data from any sensor of arbitrary size and format. More importantly, the 3DANN-M cube can be used for general image convolutions.

The general ATR process flow is depicted in Figure 2. A frame buffer holds the image and feeds a column or row of a 64x64 subwindow to CLIC every 250 ns. The 3DANN-M network then produces 64

inner-products (each with two 4096-element vectors) every 250 ns, thus accomplishing 64 convolutions of a 256x256 image with 64x64 masks in 16 ms. The 64 analog values generated by 3DANN-M are converted to 8-bit digital values and passed along to the Point Operation Processor (POP) for data fusion. Currently, the feedback memory and POP are implemented in four Adaptive Solutions' CNAPS array processor boards (each board containing 128 SIMD processors and 32 megabytes of memory)—providing flexibility to program different point operations. In the future, a custom VLSI implementation of POP may be designed and fabricated. POP takes the output from the 3DANN-M and performs the desired target recognition functions. Command and control of VIGILANTE operations (e.g., detection/classification/tracking mode command, loading of templates, point operation functions, data recording, etc.) are done though the P6 motherboard (shown as the processor/memory block in Figure 4).

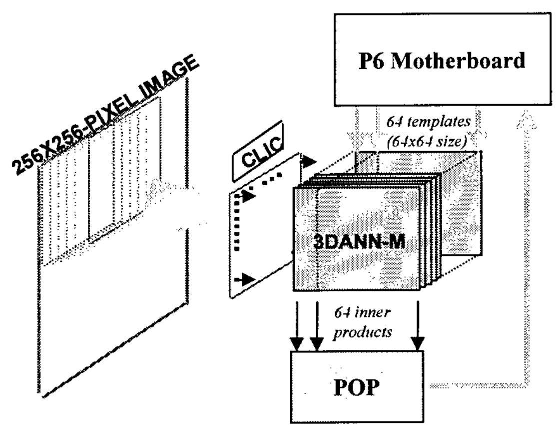


Figure 4: The VIGILANTE processing architecture that orchestrates the data flow from sensor through neural processor also serves as the basis for developing methodologies for ATR applications.

PLANNED EXPERIMENTS

A simulated CMD test scenario was selected for a series of experiments planned for VIGILANTE during 1997-1998 period. Because of the small radar cross section and unpredictable low altitude flight path of cruise missiles, new approaches to CMD other than conventional radar detection/tracking should be investigated. The unprecedented processing speed of ANTE processor and multiple spectral sensing of VIGIL will allow detection, classification, and tracking of targets like cruise missiles with great efficiency.

Flight, ground-based, and laboratory experiments will help make VIGILANTE available sooner for missile defense (NMD/TMD/CMD) applications. Since it is difficult to accurately simulate spectral responses of sensors and targets from desired vantage points, inexpensive flight experiments will provide a realistic environment for shakedown of autonomous interceptors and surveillance platforms.

Because of the ambitious scope/schedule/budget of VIGILANTE, a very lightweight helicopter (ATI Ultrasport 496) with a modified seat structure was selected as the airborne platform, see Figure 5. This compact, low-maintenance, easy to transport, and fast turnaround platform can operate both in the piloted and UAV mode for a purchase price of $200K and an operating cost of $8/hr. The unmanned mode will be employed for tests when the helicopter will be in the "line of fire" or at high altitudes (between 12,000 and 20,000 ft). These experiments will demonstrate real-time ATR from detection through precision tracking (aim-point selection).

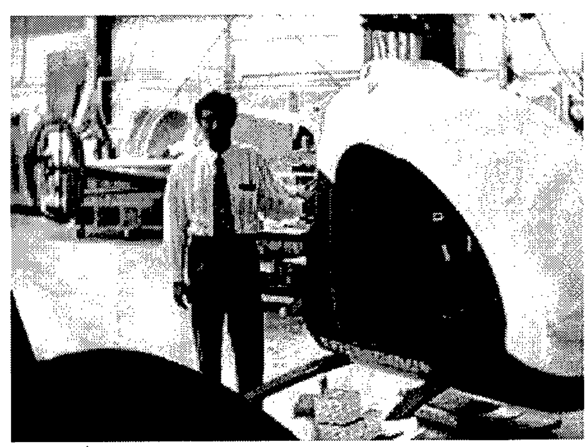
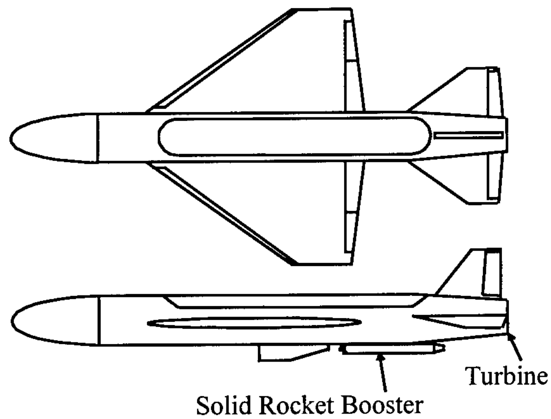


Figure 5: The VIGILANTE piloted/UAV helicopter developed by ATI is capable of 12,000 ft altitude, hovering to 98mph, 2 hrs flight time, and 500 lb load.

In addition to the ATI helicopter, a radio-controlled, reusable target, the VIGILANTE Target Vehicle (VTV)—a sub-scale model of a nominal cruise missile—serves as the cruise missile target for data collection and mock tests, see Figure 6. Although VIGILANTE piggybacks on actual cruise missile tests at NAWC, China Lake, the VTV gives

greater flexibility in terms of controlled flight paths and test scheduling.



Length	95.5 inches	(2.426 meters)
Diameter	9.7 inches	(24.64 cm)
Weight	35 pounds	

Figure 6: The VIGILANTE Target Vehicle (VTV)—8 ft long with 4 ft wing span—is a remotely piloted, turbine engine aerial vehicle capable of reaching 25,000 ft altitude and 117 mph.

Both the helicopter and the VTV are equipped with GPS receivers and transmit their position information at 10 Hz. Figure 7 depicts a typical flight profile and associated ground communication protocols planned for VIGILANTE. In September 1997, Experiment 1 (a down-looking, hovering experiment) will fly the sensor payload of IR and visible imagers with a gimbaled mirror on an active-isolation optical bench and use the GPS information to steer the gimbaled mirror and transmit image data to the ground for ANTE processing. In December, 1997, Experiment 2 will have the pilot fly the Experiment 1 sensor payload in a mock intercept flight path with the VTV as the target. Onboard sensing/processing with image-based tracking will be demonstrated for the first time in this experiment. Experiments 3 and 4, planned for 1998 will involve UAV operations of the ATI helicopter to provide closer looks at the end-game scenario and employ the next build of sensor/processor package, which will be more compact and integrated to provide simultaneous boresighted data rather than frame-dithered information.

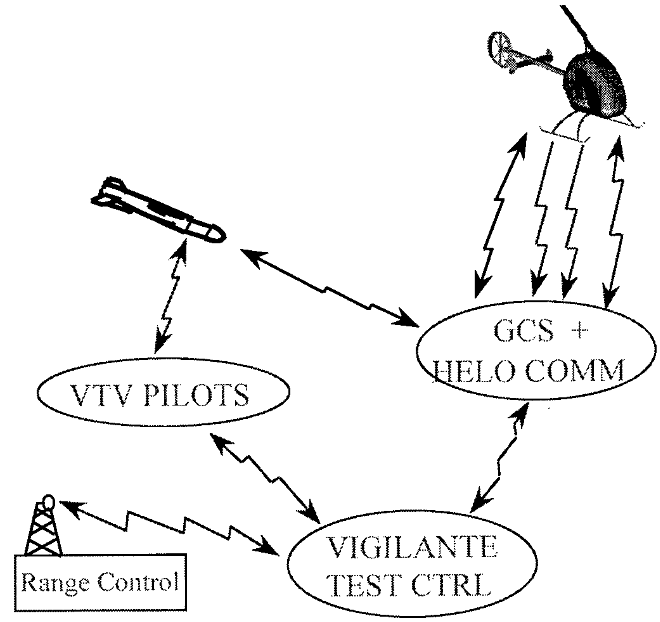
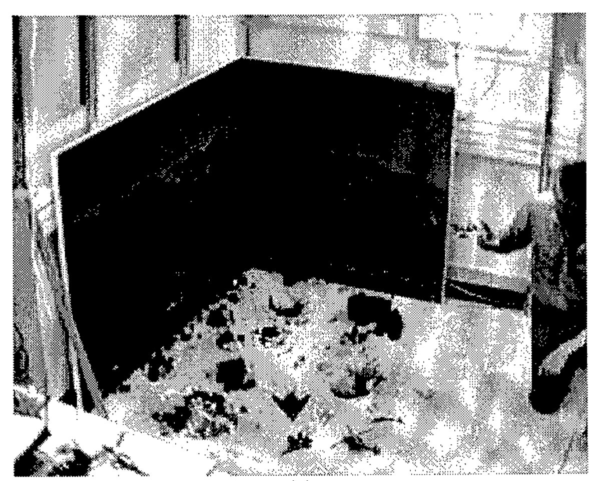
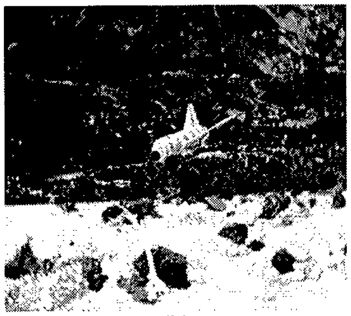


Figure 7: VIGILANTE flight experiments will demonstrate the smart sensing capability of VIGILANTE in a simulated CMD scenario with appropriate ground communication and protocols for experiment monitoring and subsequent data analysis

Lab experiments of various ATR methodologies also demonstrate the VIGILANTE smart sensing capability. We have set up a "model shop" to permit data collection of target models in many orientations, ranges, and illuminations, see Figure 8. Cluttered input scenes are produced to make the scenes more realistic.



(a)

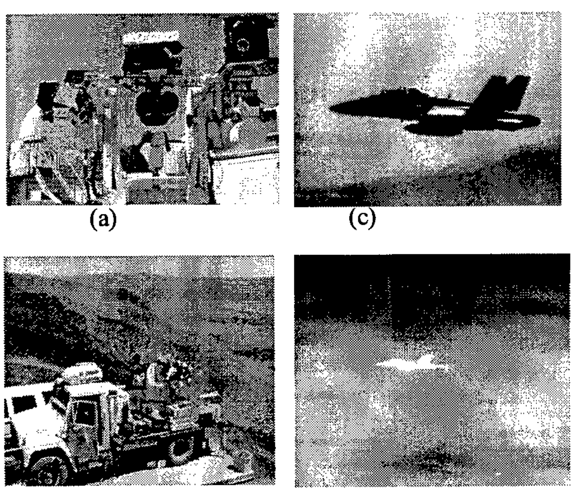


(b)

Figure 8: The VIGILANTE model shop equipped with different target models and automated image data collection system: (a) a setup to collect target imagery and (b) example input scene.

CURRENT RESULTS

Sensors: A series of field experiments to characterize the sensors were performed prior to integrating VIGILANTE sensor payload into the helicopter. Two field experiments (February and May 1997) piggy-backed on cruise missile testing at China Lake were conducted, see Figure 7. The first yielded visible (CCD camera) and IR (JPL QWIP camera) data of a fighter plane and missile. The second which has a completed setup of UV, visible, MWIR, and LWIR cameras did not provide desired target data because the cruise missile failed. A similar test setup to observe VTV is planned for September, 1997.

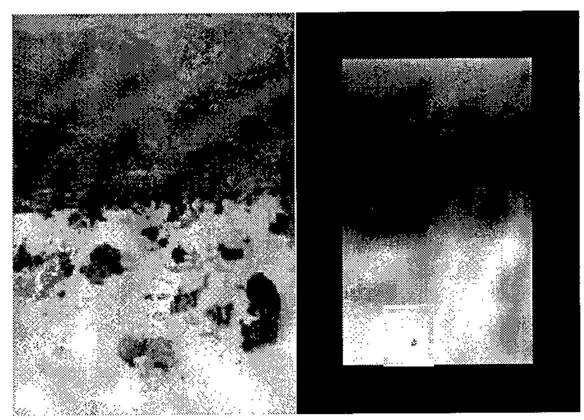


(a) (c) (b) (d)

Figure 9: Field experiment setup at China Lake: (a) UV, visible, MWIR, and LWIR cameras on M45 mount, (b) observation site at China Lake, (c) visible imagery, and (d) IR imagery.

Rich Pixel Algorithms: While efforts to integrate and test VIGILANTE are underway, progress continues in the development of ATR for real-time recognition and tracking of cruise missiles and other military targets using laboratory data. These experiments have focused on validating the design team's view of processing based upon "rich-pixel" processing, an approach which is particularly well suited to architectures similar to VIGILANTE's.

Eigenvector based hierarchical classification: Since conventional brute-force template matching is usually unreliable in highly cluttered environments (such as in Figure 10), we have investigated a hierarchical neural-network approach based on eigenvectors, see Figure 11. Figure 12 and Figure 13 show the top-level results of applying the eigenvectors of all the objects in the target library (300,000 images of helicopter, missile, and plane in various orientations). Preliminary results are promising (over 90% success rates), and it appears that, using the selected sensor/processor architecture with the 16 msec frame rate, a robust, real-time target recognition/tracking system will be realized.



TEMPLATE MASKS:

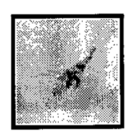


Figure 10: The brute-force template matching with a template mask of the target in a slightly different angle, although provide local maximum corresponding to the target location in the correlation image, generates a false positive result when seeking absolute maximum.

The excellent performance of the eigenvector and neural-network combination can easily be

explained in terms of having multiple-composite filter design options. In 3DANN, up to 64 composite filters can be operated in parallel on the original scene, each of which may address different discriminability/distortion-invariant characteristics. A neural-network classifier then fuses each filter output, point-by-point, thus greatly reducing the high dimensionality of nonlinear classification problems.

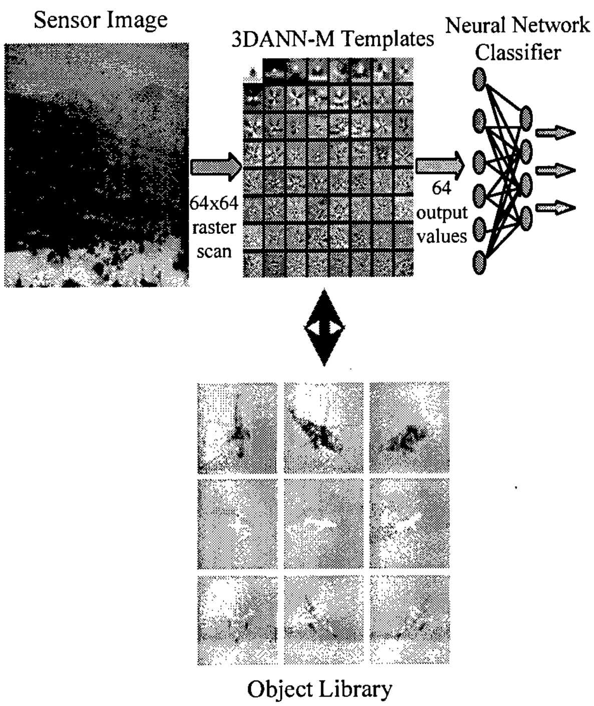


Figure 11: Hierarchical eigenvector/neural network-based target recognition synthesizes multiple composite filters using eigenvectors generated from the object library for 3DANN-M processing and classifies corresponding output value with a feedforward neural network (which can be done in POP). The hierarchy is established by returning the object library after achieving classification to reload class-specific eigenvector sets.

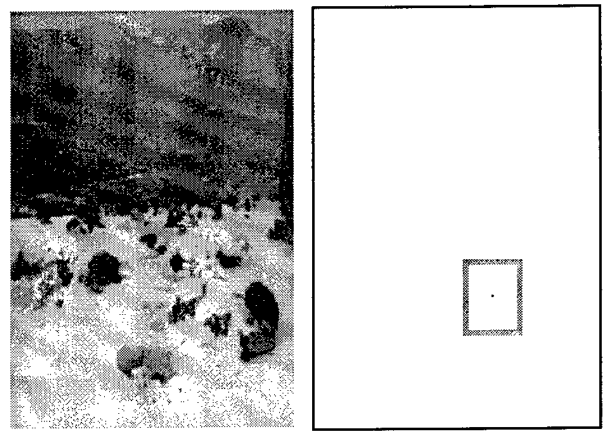


Figure 12: An example of the eigenvector/neural network output using the top-level eigenvector set generated from all the helicopter, missile, and plane images (300,000 total) in all various orientations.

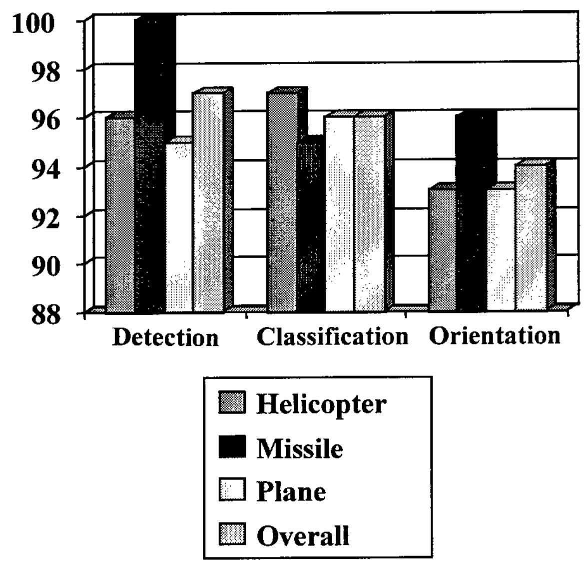


Figure 13: Performance of the eigenvector/neural network using the top-level 64-eigenvector set tested on highly cluttered scenes (5000 samples).

Hierarchical classification and tracking: The eigenvector approach permits a hierarchical recognition technique where a filter set is successively refined, permitting detailed classification of the target. The approach works as shown in *Figure 14*:

1. A very large set of general-purpose templates is used to generate 64 eigenvectors that discriminate between potential targets and nontargets. These templates are used to separate

pixels that may contain targets from pixels that don't.

2. Generate a new set of 64 eigenvectors that only discriminate among broad classes of targets to classify the features detected in step 1.
3. Use increasingly restrictive template sets to generate a set of eigenvector templates that locate the scale and orientation and more precise classification of the target.
4. Pass the results from the eigenvector matcher to the precision-tracking algorithm for precision tracking and final identification.

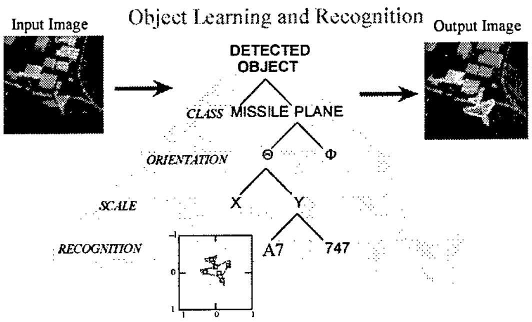


Figure 14: Hierarchical target classification and precision tracking

Precision tracking: By properly classifying the target and estimating its scale/orientation to reasonable precision, precision tracking of selected target points can be then achieved. The algorithm works by imposing geometric constraints on correlation outputs of the target's feature set (e.g., nose, wings, and tail), achieving near-zero error tracking. The algorithm works via the following steps (shown in Figure 15):

1. A series of images are generated by convolving the input image with a set of templates for detecting individual features of the object (such as the nose, left wing, etc.) The output of this step is a set of gray maps whose intensity corresponds to the likely presence of that feature (with a lot of false alarms). During the convolution operation, the images are shifted such that the features would overlay <u>**if**</u> they were present in the correct geometry, given the orientation and scale of the object.
2. The aligned images from Step 1 are fused to form a new gray map where intensity corresponds to the likelihood of a match. The example shown was fused using a winner-take-all behavior on the sum across all feature images. Neural network fusion has also been attempted.
3. The output image of Step 2 is multiplied pixel-by-pixel with the shifted convolution outputs from Step 1, resulting in new gray-maps for features which have been weighted by the likelihood that those features appear with reasonable geometry.
4. The outputs from step 3 are thresholded and shifted back out to their original locations. These feature points represent the final match.

The performance of the algorithm is shown in several cases in Figure 16.

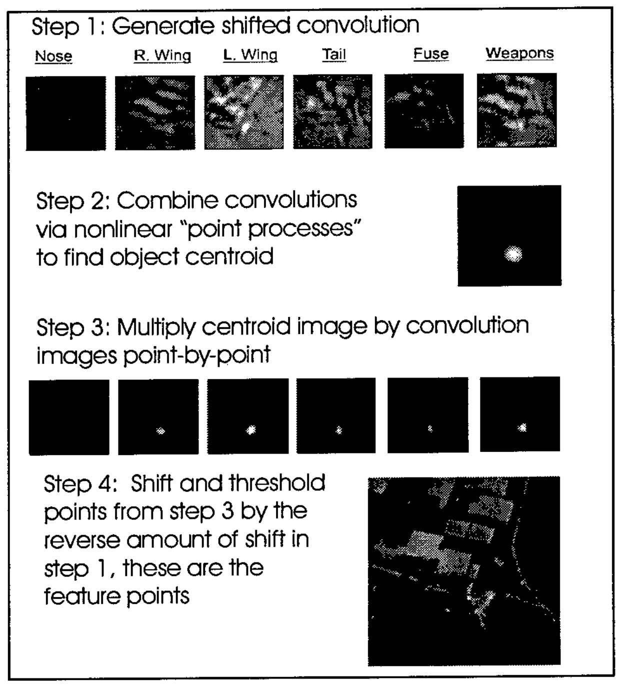
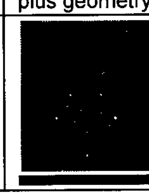


Figure 15: Precision tracking algorithm works by combining evidence, point-by-point, from shifted convolution images

	Feature detection from convolution alone	Feature detection for convolution plus geometry
Reference face		

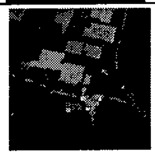
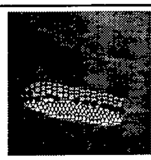
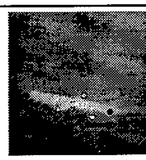
Test face		
Test image of aircraft from "shadow box"		
Test image of Tomahawk from CNN footage		

Figure 16: Performance of precision tracking algorithm on a variety of samples

CONCLUSIONS

VIGILANTE will provide an ultrafast smart sensor testbed and demonstrate an end-to-end detection/recognition/tracking in real-time through a series of flight experiments. During 1996-97, excellent progress has been made in design/development of system architecture, airborne platform and target vehicles, ATR methodologies, and field data collection. More results are anticipated in 1997-98 when the VIGILANTE payload will be integrated and its real-time/onboard smart sensing capability demonstrated in a simulated CMD scenario. VIGILANTE is a fast-pace, low-cost program providing technological breakthroughs to serve NMD and TMD needs.

ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Ballistic Missile Defense Organization through an agreement with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

REFERENCES

1. B. Bhanu, "Automatic target recognition: state of the art survey," *IEEE Trans. Aerosp. Electron. Sys.*, vol. AES-22, no. 4, pp. 364-379, 1986.
2. O. Port, "The silicon age? It's just dawning," *Business Week*, pp. 148-152, December 9, 1996.
3. D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Trans. Comp.*, vol. C-21, pp. 179-186, 1972.
4. A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theor.*, vol. 10, pp. 139-145, 1964.
5. D. Psaltis, E. G. Paek, and S. S. Venkatesh, "Optical image correlation with binary spatial light modulator," *Opt. Eng.*, vol. 23, pp. 698-704, 1984.
6. B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," *Appl. Opt.*, vol. 31, no. 23, pp. 4773-4801, 1992.
7. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 1992.
8. R. Kasturi and R. C. Jain, *Computer Vision: Principles*, IEEE Computer Society Press, Los Alamitos, California, 1991.
9. D. Hammerstrom, E. Means, M. Griffin, G. Tahara, K. Knopp, R. Pinkham, and B. Riley, "A 11 million transistor digital neural network execution engine," *Proceedings of the IEEE International Solid-State Circuits Conference,* pp. 180-181, 1991.
10. D. K. Andes, J. C. Witham, and M. D. Miles, "MAVIS: A special purpose neural computational system for ATR," *Neural Networks*, vol. 8, no. 7/8, pp. 1349-1358, 1995.
11. S. Udomkesmalee, A. Thakoor, C. Padgett, T. Daud, W.-C. Fang, and S. C. Suddarth, "VIGILANTE: An advanced sensing/processing testbed for ATR applications," *Proc. SPIE*, vol. 3069, pp. 82-93, 1997.
12. J. Carson, "On focal plane array feature extraction using a 3-D artificial neural network (3DANN)," *Proc. SPIE*, vol. 1541, Part I: pp. 141-144, Part II: pp. 227-231, 1991.
13. T. Duong, S. Kemeny, T. Daud, A. Thakoor, C. Saunders, and J. Carson, "Analog 3-D neuroprocessor for fast frame focal plane image processing," *SIMULATION*, vol. 65, no. 1, pp. 11-24, 1995.
14. C. Padgett and G. Woodward, "A hierarchical, automated target recognition algorithm for a

parallel analog processor," *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 1997.